Clustering Analysis to Evaluate Usability of Work-Flow Systems and to Monitor Proficiency of Workers

Toru Nakata

Digital Human Research Center, National Institute of Advanced Science and Technology,
Aomi 2-3-26, Koto-ku, Tokyo, 135-0064, Japan
toru-nakata@aist.go.jp

Abstract. In order to evaluate usability, we often interview the users with using questionnaire sheets. However, conventional processing methods for questionnaire data are so simple that we cannot mine maximum information from the users' opinions. This paper proposes a new method for deeper analysis of the user opinions. Using the vector quantization method, we can classify users into groups reflecting their skill grade. Also, by observing learning curves of the tasks, we can evaluate hardness of mastering each task and detect the defects of the work-flow to be improved. This paper explains the idea and mechanics of the method with referring an actual example, which is a questionnaire investigation for workers in a real office to examine usability problems on their works. The result of data analysis pointed out several very hard defects in the work-flow system, on which not only the novices but the experts are also facing difficulties. Those very hard defects cannot be solved by experience or training, because the experts cannot cope with them. Thanks to the vector quantization analysis, we can distinguish between difficult points that can be solved by workers' experience and such very hard defects that require drastic reforms for improvement from.

Keywords: Usability testing, vector quantization, questionnaire sheet method, work-flow reform.

1 Introduction

Quality of work-flow design is decisive for the efficiency of the company performance. Badly designed work-flow induces human errors, unnecessary additional costs, and accidents. The reduction of complexity of work-flow is required for various works in our life [1].

There are already analysis methods of work-flow design from the viewpoint of software engineering [2]. In such analysis, the work-flows are described as canonical form like petri net, and then their efficiencies are evaluated almost mathematically.

Clerical works, however, are often unsuitable to such formalistic analysis. Clerical works in actual offices are mainly processed by human workers. Therefore, the efficiency and comfort of the work process largely depend on usability of the tools or the system that the workers use. Usability is a psychological factor [3], [4], [5], so the

A. Marcus (Ed.): Design, User Experience, and Usability, Pt II, HCII 2011, LNCS 6770, pp. 487–496, 2011.

formalistic analysis of mathematical “work-flow analysis” is not adequate for it. Investigation focused on psychological factors is required.

One of the most difficult things to investigate psychological factors in clerical works is to monitor actual conditions of the works. Investigators of work-flow design are often afforded only a limited number of subjects for usability test, and it is difficult to survey the details of the real condition in many offices in the whole company.

In order to collect the raw and direct opinions from large number of people, we usually use the questionnaire sheet method [6], [7], [8], [9]. But it is doubtful that we have been successful in acquiring plenty information from the questionnaire data with traditional methods. The theme of this paper is how to analyze the data of the questionnaires efficiently and deeply.

The most straight-forward (and therefore most popular) method to handle the data from questionnaire sheets is calculating averages of each inquiry. The analysts warn defects of the design, which got bad reputations. When we want to get deeper information, we calculate correlations among the data and find out some hidden relations. But these analyses are elemental, and most of information contained in the data will be ignored.

Questionnaire data are generally big. The data size equals to the number of the subjects times the number of the questions. That means the data matrix contains very rich information, and the statistical analysis should be more complex to mine such big data.

The purpose of the paper is to produce an analysis method that consists of cross analysis to mine deeper information such as:

- Comprehension of users’ proficiency process,
- Categorization of users (figuring user models of novices and experts),
- Diagnose on defects by monitoring learning curves.

The paper will show the way for those purpose by referring an example of applying the vector quantization method to analyze questionnaire sheet data that is a set of workers opinions about their work-flow management.

2 Example of Analysis: Workers and Work-Flow in a Clerical Office

Since the capacity of the paper is limited, we explain the method with referring an application example in parallel.

2.1 The Worker Subjects and Their Work

A survey was carried out in a clerical office of a company. The office mainly deals financial works on papers and computers with facing many visiting customers.

The investigation using questionnaire sheet is planned to evaluate the conditions of the works and to detect real causes of difficulties and inefficiencies.

The questionnaire survey was carried out for 44 workers.

2.2 Designing Questionnaires on Usability and Difficulty of the Work-Flow

Concepts on Design and Data Processing. In general, office workers know to the last detail of the good points and defects of tools, systems, manuals, and flows of their work. So, direct questionnaire investigation for the workers is the best way to find out ways to improve the work.

The questionnaire sheets should have peripheral inquires about background and circumstance besides the main questions on opinions about usability. As shown later, peripheral data about such as personnel experiences and general opinions on working environments often have unexpected correlations toward the opinion data on usability. And then we may find hidden reasons on generation of the good points and the defects of the work-flow.

On designing questionnaire analysis method, we have to be careful about processing questionnaire data with highly complex mathematic techniques. Data of questionnaires tend to be less quantitative. The answers about opinions of the subjects are usually qualitative. To make them more quantitative, we can grade them, for instance, “good,” “neutral,” and “bad,” and deal them as numerical data. Knowing it is not accurate, we transfer the data from the ordinal style into numerical style as “1,” “2,” “3” and so on. It is just for a practical reason that numerical data are convenient for statistical analyses.

Since this enumeration from qualitative opinion data to pseudo-quantitative data is not linear transformation in general, analytical mathematic methods such as Principal Component Analysis (PCA) and Factor Analysis (FA) are not suitable. Vector quantization method is not strongly analytical, since it switches the approximation functions for each cluster unlike PCA and FA. So it may handle pseudo-quantitative data better.

The Questionnaire Sheet. We provide 49 questions asking conditions and opinions about the difficulties of certain tasks (Table 1).

The answers are replied by filling check-boxes. Each subject chooses one answer choice for each question. Each answer choice is assigned to values that are shown as numbers in the brackets as the following. By regarding these numbers as representative values of the answer choices, the answers are dealt as pseudo-quantitative numerical data.

The answer data of each subject forms a 49-dimensional vector (Fig. 1). We name the first 4 elements (about the subject’s experience) as vector L, 5th to 33rd elements as vector D (about subjective difficulties), and 33rd to 49th elements are vector T (about evaluation on the assistance tools).

We employ vector D as the key for clustering the subjects (Fig. 2). We use vector quantization method, which is one of common techniques for clustering, and then classify all the subjects into 4 groups. Each cluster are a group of the workers who gave similar opinions. We grade the groups in respect to proficiency that are evaluated from values of D (vector of difficulty).

Then we will explain the reason of generation of the proficiency differences by observing relationships among data D, L, and T. Finally, we examine the learning curves to rate the hardness of mastering each task.

Table 1. The questionnaire sheet

About experience of the tasks: Q1. Number of experienced years on the tasks: less than half year (0) / half to one year (1) / one to two years (1.5) / two or more years (3). Q2. Number of days processing the tasks in a week: rare (0) / less than one day (1) / 1-4 days (2.5) / everyday (5). Q3. Number of experienced year on reviewing colleagues' works: (answer choices are same to Q1). Q4. Reviewing frequency: (choices are same to Q2).
Recognition of overall condition of the task executing: Q5. Number of days in a week when you feel difficulty on processing the works: (choices are same to Q2). Q6. Frequency of feeling difficulty on reviewing colleagues' works: (choices are same to Q2).
Task 1: Communication with customers and colleagues: Q7. Filling up the form to represent customer's intention precisely. Q8. Handing and receiving stuffs to/from customer. Q9. Explaining circumstance of tasks to colleagues Q10. Handling and receiving stuffs to/from colleagues From Q7 to Q10, the answer choices are "(1) easy / (2) somewhat easy / (3) neutral / (4) somewhat difficult / (5) difficult."
Task 2: Filling numerical figures on the form: Q11. Do you know the most typical error pattern of the task? (1) Yes, (2) No. Q12. Can you do without the manual? Confident (1) / somewhat confident (2) / neutral (3) / somewhat unconfident (4) / unconfident (5). Q13. Do you miss filling up necessary data on the forms? Almost never (1) / rarely (2) / sometimes (3) / often (4). Q14. Can you detect colleagues' errors by reviewing? (Choices are same to Q13).
Task 3: Filling data after processing required arithmetical transformation (i.e. exchanging currency, transferring numbers of elements to number of sets, etc.) Q15-19 are same to Q11-14.
Task 4: Sealing revenue stamps. (In Japanese law, revenue stamps are basically required for receipts assigned over 30,000 JPY. Besides it, some exceptional cases exist.) Q20. Do you know the most typical error pattern of the task? (1) Yes / (2) No. Q21. Can you do without looking up the manual? (Choices are same to Q12). Q22. Do you know exceptional cases? (1) Knowing detail / (2) roughly knowing existence of problem (3) completely not knowing. Q23. Do you forget sealing necessary stamps? (Choices are same to Q13). Q24. Do you misjudge the necessity of the stamp? (Choices are same to Q13). Q25. Can you find out colleagues' errors when reviewing? (Choices are same to Q13).
Task 5: Giving and receiving signatures on forms Q26. Do you know the most typical error pattern of the task? (1) Yes / (2) No. Q27. Can you do without the manual? (Choices are same to Q12). Q28.Do you miss filling up necessary signatures? (Choices are same to Q13). Q29. Do you add unnecessary signatures mistakenly? (Choices are same to Q13).

Table 1. (*continued*)

Q30. Do you miss filling up necessary signatures in case of complex works? (Choices are same to Q13).
Q31. Do you miss filling up date or other trivial information around signatures? (Choices are same to Q13).
Q32. Do you skip verification of customers' signatures mistakenly? (Choices are same to Q13).
Q33. Can you find out colleagues' errors when reviewing (Choices are same to Q13).

Evaluation on usability of the forms and the documents
Q34. Size of letters: (1) prefer bigger / (2) keep current / (3) prefer smaller.
Q35. Spatial density: (1) prefer sparser / (2) keep current / (3) prefer thicker.
Q36. Do the forms have unnecessary boxes? (1)No / (2) some / (3) many.
Q37. Do the forms lack necessary boxes? (1)No / (2) some / (3) many.
Q38. Color of letters: (1) good / (2) neutral / (3) bad.
Q39. Do you lose pieces of the forms? (Choices are same to Q13).

Evaluation on usability of the manuals
Q40. Frequency of looking up manuals: (1) often/ (2) sometimes/ (3) rarely/ (4) never used (5) do not know the location of them.
Q41. Easy to look up? (1) Yes / (2) neutral / (3) no / (4) never used.
Q42. Easy to understand terminology? (Choices are same to Q41).
Q43. Easy to find out the procedures that you have to do? (Choices are same to Q41).
Q44. Is it helpful in general? (Choices are same to Q41).

Evaluation on usability of the quick references
Q45-49 are as of Q40-44.

2.3 Analysis of Clusters of the Workers

Correlations among Opinions of the Inquires. Tracing changes of answers of each subject (as drawn lines in Fig. 1), we can find out correlations among the questions. In contrast, we cannot tell such detail structures by just observing averages as in traditional simple analyses.

Clustering Result. Applying the clustering, we found 4 groups of the workers. The meaning of the group can be detected from the parameters concerning their experience (Table 1). We can see the detail differences of data on each questions by grouping the answers in respect to the clusters (Fig. 3). Some questions got similar opinions from all of the worker clusters, and other questions did not. Those questions with diversity of answers means they are concerned with experience of the work.

The dots mean answers of the subjects, the lines connecting the dots represent answers of same subject. There is positive correlation between opinions of inquiry Qi and Qj. Negative correlation exists between opinions of inquiry Qj and Qk. The difference of opinions allows cluster the subjects into groups Cluster 1 and 2. Taking averages of answers for each inquires (shown as broken lines in the middle) is a typical traditional method, but it neither detects such correlations nor clusters.

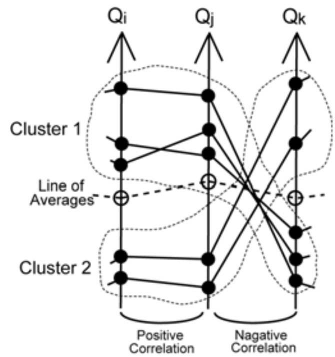

Fig. 1. Tracing each subject over the data to observe correlations

Table 2. Clusters of the workers classified by contents of the opinions

Cluster ID	Relationship to amount of experiment	Number of workers
1 ★	Workers with long experiences on the job.	12
2 ▼	Workers doing the works very frequently.	15
3 ●	Intermediate workers.	10
4 ▲	Novice workers.	7

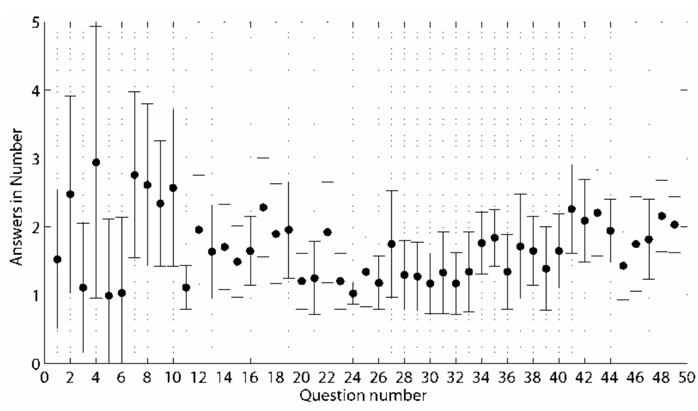

Fig. 2. Means and standard deviations of the answers over all subjects. The answers are transferred to pseudo-quantitative values.

Reasoning the Clusters. When tracing each subject (like shown in Fig. 1), we may find the typical patterns of the answers. By using clustering methods such as vector quantization we are able to nominate the patterns precisely.

In questionnaire sheet investigation about usability, there exist several typical answer patterns with keeping clear separation gaps to other patterns. This tendency appears because the population of the subjects is consisted of different groups, i.e. novices and experts. Difference of proficiency of the workers makes differences on the opinion answers. However, other factors besides the experience length may have influences.

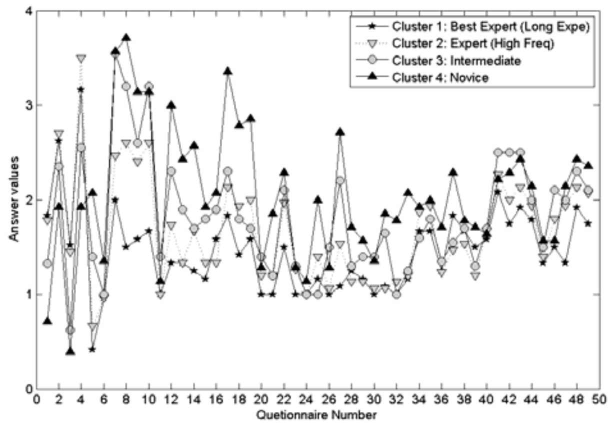

Fig. 3. Average of answers of each cluster

To detect what factors are effective and to check their power, we process the data as 3 steps:

1. Select the parameters which are candidates of cause factors of the cluster separation. (For example, the data about background of each worker should be dealt as a candidate.)
2. Do clustering of the data without using the candidate parameters. Make clusters in respect to only the rest part of the answers, i.e. data about opinions about the work-flow.
3. Monitor the correlation between the candidate parameter and difference of the clusters, by tracing answers of each cluster (Fig. 4). If a set of the trace lines from a cluster to a certain candidate parameter converges in unique position, it means that the cluster has strong relationship to the candidate parameter. Narrowness of convergence and separation against other clusters indicate the power of the relationship. Then we can evaluate the candidate parameters.

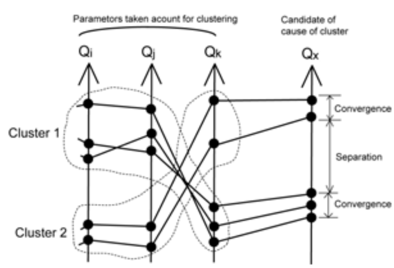

Fig. 4. Detecting the cause of the cluster separation. The cause parameter has strong correlations to difference of the cluster. By tracing the data of each subject (drawn as lines), we can observe the power of influence to the cluster separation.

2.4 Difficulties of the Work and Learning Curves

Concept of Learning Curve Analysis. We can find defects in the work-flow by observing of "learning curves." Learning curves can be drawn by plotting opinion answers about the difficulties of each task in the work-flow in respect to the length of experience which each worker has.

Shape of learning curve reflects condition of education and training for the workers about the task (Fig. 5). Bent learning curves imply mature of proficiency is fast or slow. Slow mature curve indicates that the task is very complex, so that only the experts can do the task. We can find that such difficult tasks must have certain problems on their process and should be reviewed and fixed.

Strange curves may be appears. Normally, the more you have experience, the less you make mistakes. Such monotonic decreasing relationships between experience and hardness of the work are common to normal works. But non-monotonic learning curves sometimes appear. They mean the tasks are suffering disturbance than are caused by certain strange circumstances of the work-flow.

Such strange curves, however, are not rare in actual investigations. There may be several reasons for generation of non-monotonic, and the most plausible reason is that the expert workers are often entrusted with the most difficult tasks. So over-all hardness of the work is stronger than that for the novices and intermediate workers. When the strange curves are found, we should review assistances and special training for over-intermediate workers to process their complex works.

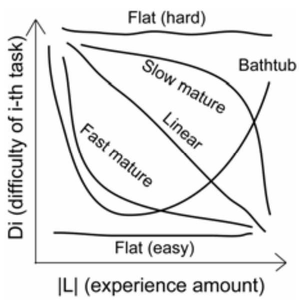

Fig. 5. Some typical patterns of learning curves

Results and Discussion on Learning Curves. We got the learning curves shown in Fig. 6.Non-monotonic curves are found in Q11, 19, 20, 25, 26, 34, 37, 41, and 42. The working conditions concerning thoses questions are thought having special and severe prombles. For instance, non-monotonic learning curves appeared in questions about knowleges on occurrence distrubution of typical clerical mistakes (Q11, Q20, and Q26) and reviewing skill (Q19 and Q25). This means that the experts are facing more difficult tasks and are less confident than the intermediate workers, so that there should be trainings for the experts to improve their skill.

Fig. 6. Learning curves from the novice cluster (▲), the intermediate cluseter (●) to the two expert clusters (★ and ▼). X-axis: experience amount estimated as norm of experience vector L. Y-axis: average of the answers in each cluster. For inquries Q5 to Q49, the larger the value is, the worse the opinions about the inqurie are.

References

1. OECD: Cutting Red Tape — National Strategies for Administrative Simplification. OECD (2006)
2. Wil, v.d.A., Hee, v.K.: Workflow Management. MIT Press, Cambridge (2002)
3. Gilb, T., Weinberg, G.M.: Techniques for Reliable Keyed Input. Winthrop Pub. (1976)
4. Gillespie, R.: Manufacturing knowledge — A history of the Hawthorne experiments. Cambridge Univ. Press, Cambridge (1991)

5. Freiberg, K., Freiberg, J.: Nuts! — Southwest Airlines Crazy Recipe for Business and Personal Success. Bard Press (1996)
6. Lewis, J.R.: IBM Computer Usability Satisfaction Questionnaires: Psychometric Evaluation and Instructions for Use. IBM Technical Report 54.786 (1993)
7. Lin, H.X., Choong, Y., Salvendy, G.: A Proposed Index of Usability: A Method for Comparing the Relative Usability of Different Software Systems. Behaviour & Information Technology 16(4/5), 267–278 (1997)
8. Kirakowski, J., Cierlik, B.: Measuring the Usability of Web Sites. Human Factors and Ergonomics Soc. Annual Meeting Proc., Communications (5), 424–428 (1998), Human Factors and Ergonomics Soc.
9. Gersho, G.A., Robert, M.: Vector Quantization and Signal Processing. Kluwer Academic Pub., Dordrecht (1992)